Identification of *AKARI* infrared sources by the Deep HSC Optical Survey: construction of a new band-merged catalogue in the North Ecliptic Pole Wide field

Journal Item

page.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **500,** 4078–4094 (2021) doi:10.1093/mnras/staa3359

Identification of *AKARI* infrared sources by the Deep HSC Optical Survey: construction of a new band-merged catalogue in the North Ecliptic Pole Wide field

Seong Jin Kim [1]★ Nagisa Oi [2] Tomotsugu Goto,[1] Hiroyuki Ikeda,[3,4] Simon C.-C. Ho [1] Hyunjin Shim [5] Yoshiki Toba [6,7,8] Ho Seong Hwang [9] Tetsuya Hashimoto [1,10] Laia Barrufet,[11,12] Matthew Malkan,[13] Helen K. Kim [13] Ting-Chi Huang [14,15] Hideo Matsuhara,[15] Takamitsu Miyaji,[16]† Chris Pearson [12,17,18] Stephen Serjeant,[17] Daryl Joe D. Santos,[1] Eunbin Kim [9] Agnieszka Pollo,[19,20] Woong-Seob Jeong,[9] Ting-Wen Wang [1] Rieko Momose [21] and Toshinobu Takagi[22]

Affiliations are listed at the end of the paper

Accepted 2020 October 23. Received 2020 October 23; in original form 2020 August 31

ABSTRACT

The North Ecliptic Pole field is a natural deep-field location for many satellite observations. It has been targeted many times since it was surveyed by the *AKARI* space telescope with its unique wavelength coverage from the near- to mid-infrared (mid-IR). Many follow-up observations have been carried out, making this field one of the most frequently observed areas with a variety of facilities, accumulating abundant panchromatic data from the X-ray to the radio wavelength range. Recently, a deep optical survey with the Hyper Suprime-Cam (HSC) at the Subaru telescope covered the NEP-Wide (NEPW) field, which enabled us to identify faint sources in the near- and mid-IR bands, and to improve the photometric redshift (photo-z) estimation. In this work, we present newly identified *AKARI* sources by the HSC survey, along with multiband photometry for 91 861 *AKARI* sources observed over the NEPW field. We release a new band-merged catalogue combining various photometric data from the *GALEX* UV to submillimetre (sub-mm) bands (e.g. *Herschel*/SPIRE, JCMT/SCUBA-2). About ∼20 000 *AKARI* sources are newly matched to the HSC data, most of which seem to be faint galaxies in the near- to mid-infrared *AKARI* bands. This catalogue is motivating a variety of current research, and will be increasingly useful as recently launched (*eROSITA*/*ART-XC*) and future space missions (such as *JWST*, *Euclid*, and *SPHEREx*) plan to take deep observations in the NEP field.

Key words: catalogues – galaxies: evolution – cosmology: observations – infrared: galaxies.

1 INTRODUCTION

To answer questions concerning cosmic star-formation history and galaxy evolution, it is critical to have a comprehensive understanding of the infrared (IR) luminous populations of galaxies (Kirkpatrick et al. 2012; Casey, Narayanan & Cooray 2014; Madau & Dickinson 2014; Sanders et al. 2014). These are presumably star-forming systems containing a large amount of dust, where critical phases in their activities of star formation (SF) or active galactic nuclei (AGNs) take place, hidden behind the obscuring dust (Goto et al. 2010; Lutz 2014; Galliano, Galametz & Jones 2018; Hickox & Alexander 2018).

Wide-field cosmological surveys at IR wavelengths are the most efficient way to collect data for various populations of galaxies, especially for dusty star-forming galaxies (dusty SFGs, DSFGs) and obscured AGNs, at different cosmological epochs (Matsuhara et al. 2006; Hwang, Serjeant & Lee 2007; Hwang et al. 2010; Toba et al. 2015). Statistically significant samples of dusty galaxies based on large-area surveys covering significant cosmological volumes have to be obtained. Also, follow-up surveys should be made to sample spectral energy distributions (SEDs): a comprehensive physical description requires wide wavelength coverage to capture the range of processes involved. Most importantly, a deep optical follow-up survey is necessary because optical identification is an essential prerequisite to understand the nature of the sources (Sutherland & Saunders 1992; Hwang, Geller & Kurtz 2012), e.g. star–galaxy separation, or to derive the photometric redshift (photo-z).

The North Ecliptic Pole (NEP; RA = $18^{h}00^{m}00^{s}$, Dec. = $66^{\circ}33'88''$) has been a good target for deep, unbiased, and contiguous surveys for extra-galactic objects such as galaxies, galaxy clusters, and AGNs because the NEP is a natural deep-field location for a wide class of observatory missions (Serjeant et al. 2012). Many astronomical satellites, such as *ROSAT* (Henry et al. 2006), *GALEX*[1] (Burgarella et al. 2019), and the *Spitzer Space Telescope*[2] (Jarrett et al. 2011; Nayyeri et al. 2018), have accumulated a large number of exposures towards the NEP area because the Earth-orbiting satellites

★ E-mail: seongini@gmail.com
† Present address: Leibniz Institut für Astrophysik Potsdam (AIP), An der Sternwarte, Potsdam D-14482, Germany.

[1] http://www.galex.caltech.edu/
[2] http://www.spitzer.caltech.edu/

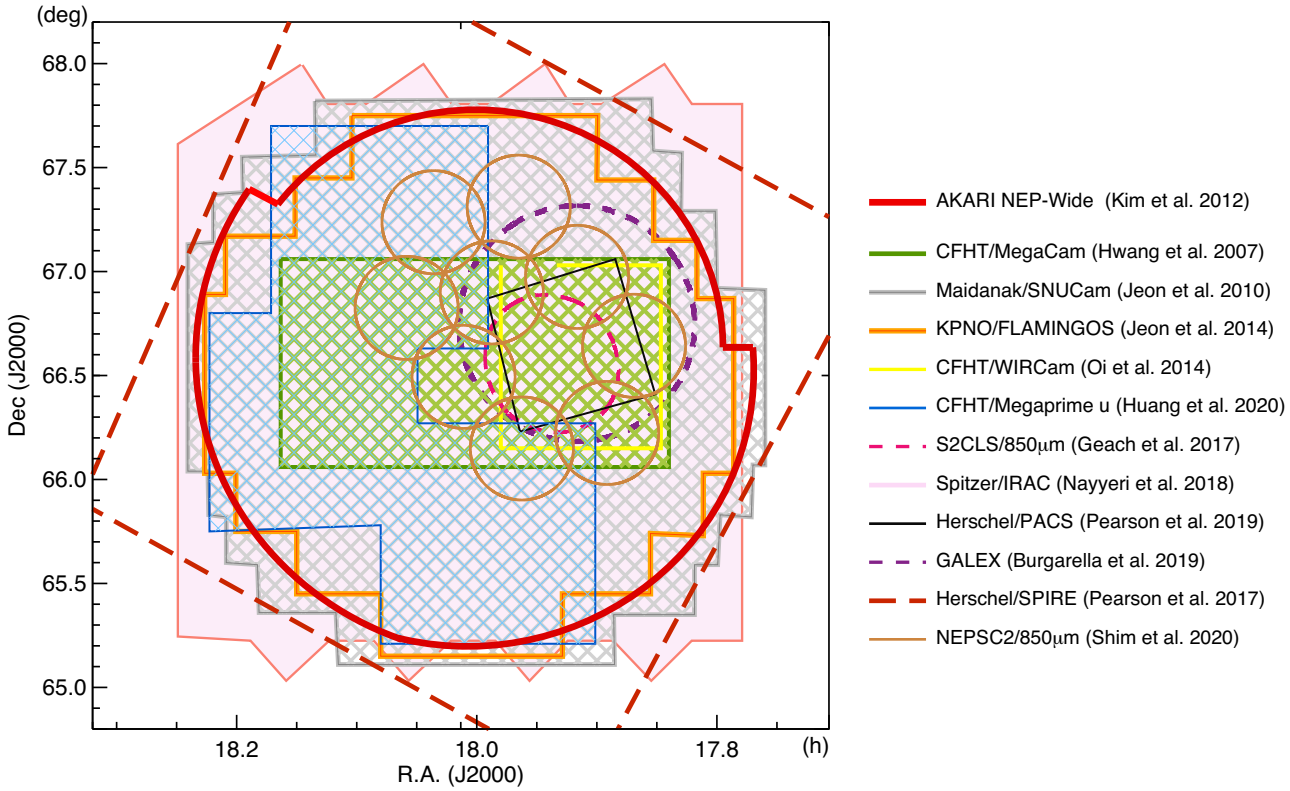

Figure 1. An overall map showing a variety of surveys around the NEP. The red circular area shows *AKARI*'s NEP-Wide field (Kim et al. 2012). The green and grey (hatched) areas represent the optical surveys done with the CFHT MegaCam (Hwang et al. 2007) and the Maidanak SNUCam (Jeon et al. 2010), respectively. The yellow square shows a slightly deeper observation with the MegaCam as well as the WIRCam on the NEP-Deep field (Oi et al. 2014). The area surrounded by the blue line shows the additional *u*-band observation by the MegaPrime (Huang et al. 2020). The pink shaded area indicates the recent near-IR survey with *Spitzer* (Nayyeri et al. 2018). The small black square inside the yellow box shows the area observed by *Herschel*/PACS (Pearson et al. 2019). The broken magenta circle overlaid with the black square indicates the area observed by S2CLS (Geach et al. 2017). The nine brown circles around S2CLS show the areas surveyed by the NEPSC2 850 μm mapping programme with SCUBA-2 (Shim et al. 2020). The largest rhombus (brown long-dashed line) shows the *Herschel*/SPIRE coverage (Pearson et al. 2017).

pass over the ecliptic poles and, for the Earth-trailing satellites, these poles are always in the continuous viewing zone. *AKARI* (Murakami et al. 2007) also devoted a large amount of observing time (a span of more than a year) to covering a wide area over the NEP using the infrared camera (IRC, Onaka et al. 2007) with excellent visibility thanks to its polar Sun-synchronous orbit (Matsuhara et al. 2006).

A noticeable aspect of this NEP field surveyed by *AKARI* is that ancillary optical data sets are abundant, supporting the identification of the infrared sources and thus enabling proper subsequent analysis (Kim et al. 2012). In addition, many other surveys or follow-up observations of the NEP (but on a limited area) have been carried out from the X-ray to radio wavelengths to cover the NEP area (White et al. 2010, 2017; Krumpe et al. 2015; Geach et al. 2017; Burgarella et al. 2019; Pearson et al. 2019) since *AKARI* obtained valuable data sets (see Fig. 1). However, a fraction ($\sim$ 30 per cent at N4) of the IR sources detected by *AKARI* have been left unidentified by optical data because of the insufficient depths and incomplete areal coverage of the previous optical surveys. The different photometric filter systems used in different surveys also hampered homogeneous analyses based on unbiased sample selection; therefore a deeper and more coherent optical survey on this field was necessarily required.

A new deep optical survey consistently covering the entire NEP-Wide (NEPW) field was carried out by the Hyper Suprime-Cam (HSC, Miyazaki et al. 2018) with five photometric filter bands (*g*, *r*, *i*, *z*, and *y*). These HSC data were reduced (Oi et al. 2020) by the recent version of the pipeline (v6.5.3, Bosch et al. 2018), which allowed the depth of the new optical data to reach $\sim$2 mag deeper (at the *g* and *i* bands) than the previous optical survey with the Canada–France–Hawaii Telescope (CFHT, Hwang et al.). In addition, supplemental observation using the CFHT/MegaPrime (Huang et al. 2020) replenished the insufficient coverage of the u^* band from the previous CFHT surveys, which brings photo-*z* accuracy improvement along with these new HSC data (Ho et al. 2020). The source-matching and band-merging processes (see Section 2 for details) have been encouraging various subsequent works such as the recent luminosity function (LF) update (Goto et al. 2019), properties of mid-IR (MIR) galaxies detected at 250 μm (Kim et al. 2019), estimation of the number fraction of AGN populations (Chiang et al. 2019), a study on high-*z* populations (Barrufet et al. 2020), obscured AGN activity (Wang et al. 2020), merger fractions depending on star-formation mode (Kim et al., in preparation), AGN activities depending on the environment (Santos et al., in preparation), machine-learning algorithms to classify/separate IR sources (Poliszczuk et al., in preparation; Chen et al., in preparation), and cluster-candidate finding (Huang et al., in preparation), even on the *AKARI* sources without any HSC counterpart (Toba et al. 2020). The science on the NEP initiated by *AKARI* is now entering a new era with momentum driven by Subaru/HSC observations as well as current survey projects, such as a homogeneous spectroscopic survey (MMT2020A/B, PI: H. S. Hwang) and 850 μm mapping over the

entire NEP area using the Submillimetre Common-User Bolometric Array 2 (SCUBA-2) at the James Clerk Maxwell Telescope (Shim et al. 2020). More extensive imaging observations with HSC are still ongoing with spectroscopy with Keck/DEIMOS+MOSFIRE as part of the Hawaii Two-O project (H20).[3] *Spitzer* also finished its ultradeep NIR observations recently as one of the legacy surveys (PI: Capak) before it retired in early 2020 to carry out a precursor survey for *Euclid* (Laureijs et al. 2011), the *James Webb Space Telescope* (*JWST*, Gardner et al. 2006), and the *Wide-Field Infrared Survey Telescope* (*WFIRST*, Spergel et al. 2015) over this field. *Spektr-RG* was launched in 2019 to the L2 point of the Sun–Earth system (1.5 million km away from us) and *eROSITA* (Merloni et al. 2012) started its mission towards the NEP. The Spectro-Photometer for the History of the Universe, Epoch of Reinonization, and Ice Explorer (*SPHEREx*, Doré et al. 2016, 2018) mission is also planning to target this field.

The main goal of this work is to identify optical counterparts of the *AKARI*/NEPW sources with more reliable optical photometry of the HSC images (even for the faint NIR sources), and cross-check with all available supplementary data covering this field to build panchromatic data. We briefly describe various data supporting the *AKARI*/NEPW data, but mostly focus on explaining how we matched sources and combined all the data together. This paper is organized as follows. Section 2 introduces the HSC and *AKARI* data, and describes the detailed process for cross-matching the sources between them. In Section 3, we present the complementary data sets used to construct the multiband catalogue. We describe their optical–IR properties (in statistical ways) in Section 4. Section 5 gives the summary and conclusions. All magnitudes are presented in the AB magnitude system.

2 IDENTIFICATION OF *AKARI*'S NEP-WIDE SOURCES USING DEEP HSC DATA

2.1 *AKARI* NEP-Wide survey data

The NEP-Wide (NEPW) survey (Matsuhara et al. 2006; Lee et al. 2009; Kim et al. 2012), as one of the large-area survey missions of the *AKARI* space telescope (Murakami et al. 2007), has provided us with a unique IR data set, sampling the near- to mid-IR wavelength range without large gaps between the filter bands (the circular area surrounded by a thick red line in Fig. 1). In this programme, they observed a 5.4 deg^2 circular area centred at the NEP using nine photometric bands covering the range from 2–25 μm continuously. The overall strategy of the survey was explained by Lee et al. (2009). Kim et al. (2012) presented the description of the data reduction, source detection, photometry, and catalogue. They also combined the nine separate catalogues (i.e. for N2, N3, N4 in the NIR, S7, S9W, S11 in the MIR-S, and L15, L18W, L24 in the MIR-L channels) along with the optical identification/photometry. Before combining, they carefully discriminated the spurious objects and false detection, in order to confirm the validity of the IR sources: they tried to identify optical counterparts using the CFHT (Hwang et al. 2007) and Maidanak (Jeon et al. 2010) data and then cross-checked against the NIR *J*-, *K*-band data obtained from the Kitt Peak National Observatory (KPNO)/FLAMINGOS (Jeon et al. 2014).

The numbers of sources detected at the nine IRC bands (`DETECT_THRESH=3`, `DETECT_MINAREA=5`) by SExtractor (Bertin & Arnouts 1996) are 87 858, 104 170, 96 159, 15 390, 18 772, 15 680, 13 148, 15 154, and 4019 (the detection limits are 21 mag in NIR, 19.5–19 mag in MIR-S, and 18.6–17.8 mag in MIR-L), respectively (Kim et al. 2012). A significant fraction of these sources (17 per cent of the N2, 26 per cent of the N3, and 30 per cent of the N4 sources) did not have optical data (mostly because they are not detected in optical surveys). In addition, ∼4 per cent of the NIR sources were finally rejected because they were detected at only one NIR band (e.g. N2, N3, or N4) and are strongly suspected to be false objects: they suffered from 'multiplexer bleeding trails' (MUX-bleeding) due to the characteristics of the InSb detector array (Holloway 1986; Offenberg et al. 2001) so that source detection near the MUX-bleeding was strongly affected by artificial effects and spurious objects near the abnormally bright pixels. Also, the false detection caused by cosmic ray hits was serious at the N4 band, mostly because the telescope time for taking dithering frames was sometimes assigned to telescope manoeuvring (by the IRC03 observation template). If a certain source were detected at only one NIR band, it could potentially be an artefact or false detection. Therefore, the sources detected at only one NIR band were excluded in the first release of the band-merged NEPW catalogue. In the MIR bands, cosmic ray hits and other artefacts are not numerous, and so the sources detected by only one MIR band were included.

Note that data from the *AKARI* NEP-Deep (NEPD) survey (Wada et al. 2008; Takagi et al. 2012), which is similar to the NEPW survey (with consistent photometries and the same astrometric accuracy) but different in terms of the coverage (0.7 deg^2) and the depth (∼0.5 mag deeper), are not included in this work.

We expect that the new optical data obtained by the HSC (Oi et al. 2020) will allow us to identify more IR sources, most of which are faint in the IR as well as the optical bands. Also, we may be able to examine if there are any real NIR sources that have been rejected just because they did not have any counterpart in the other bands (from the optical to MIR). We therefore repeated the merging of nine *AKARI* bands without any exclusion process, in order to attempt to recover possibly real *AKARI* sources not included in the study by Kim et al. (2012). Sources detected by at least one *AKARI* band can be included in the new *AKARI* nine-band-merged catalogue. Spurious objects or artefacts can be excluded later if we find any, during further analysis.

When we carried out this procedure, we began with the matching between the N2 and N3 bands first. After that, we used these results against the N4 band using N3 coordinates. In cases without N3 coordinates (i.e. a source detected at N2 but not detected at N3), we took N2 coordinates for the matching against N4. This process went through down to L18W or L24. In the resulting catalogue, we kept the coordinates from the shortest and longest bands in this matching process. Therefore, if a certain source was detected neither at N2 nor L24 but at the other bands, then the coordinate information for the shortest band is from N3 and the longest from L18W. There is no systematic offset between the astrometry from different *AKARI* bands (all of them are fixed at the same focal plane). We eventually registered 130 150 IR sources in the new NEPW catalogue, which were detected at least one of the *AKARI* bands from N2 to L24.

2.2 Deep optical survey with HSC over the NEP

A deep optical survey covering the whole NEP field with the HSC was proposed (Goto et al. 2017) in order to detect all the IR sources observed by *AKARI*. Two optical data sets, one obtained with the CFHT (the central green square in Fig. 1; Hwang et al. 2007) and one with the Maidanak telescope (the grey area in Fig. 1; Jeon et al. 2010), have been used to support the *AKARI* IR data. However, the depths (5σ) of these optical surveys (25.4 and 23.1 AB mag at the r'

[3] https://project.ifa.hawaii.edu/h20

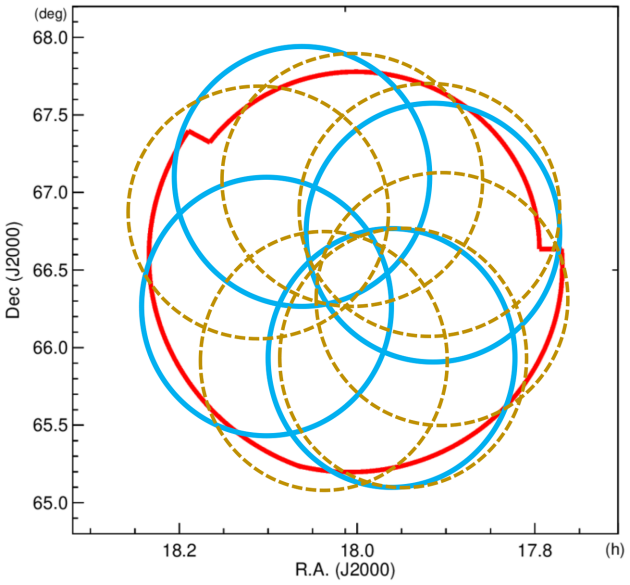

Figure 2. The map showing the HSC coverage over the NEPW survey area (red circular area; Kim et al. 2012). The areas marked by yellow dashed circles show the region observed by the HSC *r* band, while the four blue solid circles indicate the region observed by the *g*, *i*, *z*, and *y* bands (Oi et al. 2020).

and *R* bands, respectively) were insufficient to identify all the *AKARI* IR sources. A slightly deeper observation by MegaCam (Oi et al. 2014) on a smaller field was also carried out, but the areal coverage (0.67 deg^2) was only for the NEPD field (the yellow box in Fig. 1).

Goto et al. (2017) intended to use the large field of view (FoV; 1.5 deg in diameter, see Fig. 2) of the HSC so that the entire NEPW field of *AKARI* could be covered by taking only 4 FoVs (for the *g*, *i*, *z*, and *y* bands; the blue circles in Fig. 2). 10 FoVs in total were allocated in six nights (Goto et al. 2017; Oi et al. 2020), including the *r* band, to take the whole NEPW field using those five HSC filter bands. The *r*-band imaging was taken earlier during the first observation in 2014, where the observations suffered from air disturbance (including a dome shutter opening error), which made the seeing at the *r* band worse (1.25 arcsec) than that at the other four (0.7–0.8 arcsec) obtained later during the second observations (2015 August) (Oi et al. 2020).

The data reduction was carried out with the official HSC pipeline, `hscPipe` 6.5.3 (Bosch et al. 2018). Apart from the fundamental pre-processing procedures (e.g. bias, dark, flat, fringe, etc.), the performance in the sky subtraction and artefact rejection was enhanced in this recent version. In particular, the peak-culling procedure was included to cull spurious or unreliable detections, which improved the source detection results in the pipeline process. Owing to the bad seeing (∼1.25 arcsec) in the *r* band, the source detection was carried out on the *gizy* band stacked image. However, the photometry was forced to be performed at all five bands. All these procedures with recent pipeline (e.g. updated flag information of the `hscPipe` 6.5.3, Bosch et al. 2018) eventually helped to resolve the issues that have been reported for a few years regarding false detections (e.g. damaged sources near the boundary or along the image edge of each frame, and effects from the saturation of bright stars, etc.). The 5σ detection limits are 28.6, 27.3, 26.7, 26.0, and 25.6 mag at *g*, *r*, *i*, *z*, and *y*, respectively. The limiting magnitudes of the *g*, *r*, *i*, and *z* bands of the previous CFHT data were 26.1, 25.6, 24.8, and 24.0 mag, respectively. We therefore obtained 1.7–2.5 mag deeper optical data at the corresponding filter bands (even though the effective wavelengths of the filters are slightly different; see Table 1). Finally, we catalogued 3.25 million sources observed by the optical survey with the HSC along with a large number of parameters appended from the HSC data pipeline.

The magnitude at each band was given in terms of `Cmodel` photometry, which performs well for galaxies and asymptotically approaches point spread function (PSF) photometry for compact sources. The colours estimated with `Cmodel flux`[4] are robust against seeing conditions (Huang et al. 2018). As we mentioned in the previous paragraph, the seeing conditions for the *r* band are different to those for the other four bands; therefore, by taking `Cmodel flux` to calculate colours, these different seeing condition effects can be compensated.

2.3 Matching of the *AKARI* infrared sources against the HSC optical data

After *AKARI* band merging (Section 2.1), we performed source matching between the *AKARI* and HSC data. To identify the counterparts to each other by positional matching, a reasonable search radius was assigned. Fig. 3 summarizes how we decided on the radius for the source matching between the *AKARI* and HSC data.

Fig. 3(a) shows the distribution of the positional offsets of the matched sources on the ΔRA versus ΔDec. plane (before deciding on the radius, we used 3 arcsec as a tentative criterion). The contours in the reddish dotted lines with numbers represent the number densities of the green dots normalized by the peak value at the centre. A yellow circle indicates the 3σ radius determined based on the Gaussian (magenta curves) fitted to the histograms in panels (b) and (c). Here, the 3σ radius corresponds to 1.78 arcsec, where the source density in the ΔRA cos (dec) versus ΔDec. plane approaches the 1 per cent level compared to the density peak.

Within this matching radius, we have 111 535 *AKARI* sources matched against the HSC optical data, which were finally divided into two groups: clean (91 861) versus flagged (19 674) sources based on the HSC flag information such as `base_PixelFlags_flag_bad` (to discriminate bad pixels in the source footprint), `base_PixelFlags_flag_edge` (to indicate that a source is in the masked edge region or has no data), and `base_PixelFlags_flag_saturatedCenter` (to note that saturated pixels are in the centre of a source). These parameters helped us when we discriminated unreliable results with saturated sources or those lying at the image edge/border. In this work, we construct a band-merged catalogue for the 'clean' sources only, excluding the flagged ones because the derivation of photo-*z* or physical modelling by spectral energy distribution (SED) fitting requires accurate photometry.

The remaining 23 620 sources did not match any HSC data (non-HSC, hereafter),;some of wthese eem to be obscured objects in the optical bands, residing in the high-*z* (>1) universe (Toba et al. 2020).

The histogram in Fig. 3(d) shows the distribution of the matched sources as a function of radius interval. The green (red) bars show the number of clean (flagged) sources matched in each radius bin with a 0.5 arcsec width. A yellow mark indicates the matching radius (1.78 arcsec; therefore all the sources in the 1.5–2.0 bin are matched within this radius). The green histogram shows that half of the clean sources (48 per cent, 43 789) are matched within 0.5 arcsec positional offsets and 34 per cent (31 215) are matched with offsets between 0.5 and 1.0 arcsec. Therefore, within a 1 arcsec radius, we have

[4]AB magnitude $=-2.5\log_{10}$(`Cmodel flux`) + 27.

Table 1. Summary of the multiwavelength data sets: the detection limits.

Data	Band	Effective wavelength (μm)	(5σ) detection limit AB/μJy
	N2	2.3	20.9/15.4
	N3	3.2	21.1/13.3
AKARI/IRC	N4	4.1	21.1/13.6
NEP-Wide survey	S7	7	19.5/58.6
5.4 deg^2	S9W	9	19.3/67.3
(Kim et al. 2012)	S11	11	19.0/93.8
	L15	15	18.6/133
	L18W	18	18.7/120
	L24	24	17.8/274
	g	0.47	28.6/0.01
Subaru/HSC	*r*	0.61	27.3/0.04
5.4 deg^2	*i*	0.76	26.7/0.08
(Oi et al. 2020)	*z*	0.89	26.0/0.14
	y	0.99	25.6/0.21
CFHT/MegaPrime	*u*	0.36	25.4/0.25
3.6 deg^2 (Huang et al. 2020)			
	u^*	0.39	26.0/0.16
CFHT/MegaCam[a]	*g*	0.48	26.1/0.13
2 deg^2 (Hwang et al. 2007)	*r*	0.62	25.6/0.21
0.7 deg^2 (Oi et al. 2014)	*i*	0.75	24.8/0.39
	z	0.88	24.0/0.91
Maidanak/SNUCam	*B*	0.44	23.4/1.58
4 deg^2 (Jeon et al. 2010)	*R*	0.61	23.1/2.09
	I	0.85	22.3/4.36
KPNO/FLAMINGOS	*J*	1.2	21.6/8.32
5.1 deg^2 (Jeon et al. 2014)	*H*	1.6	21.3/10.96
CFHT/WIRCam	*Y*	1.02	23.4/1.58
0.7 deg^2 (Oi et al. 2014)	*J*	1.25	23.0/2.29
	K_s	2.14	22.7/3.02
Spitzer/IRAC	IRAC1	3.6	21.8/6.45
7 deg^2 (Nayyeri et al. 2018)	IRAC2	4.5	22.4/3.95
0.4 deg^2 (Jarrett et al. 2011)	IRAC3	5.8	20.3/27.0
	IRAC4	8	19.8/45.0
	W1	3.4	18.1/18
WISE	W2	4.6	17.2/23
(Jarrett et al. 2011)	W3	12	18.4/139
	W4	22	16.1/800
Herschel/PACS[b]	Green	100	14.7/4.6 mJy
0.44 deg^2 (Pearson et al. 2019)	Red	160	14.1/8.7 mJy
Herschel/SPIRE[c]	PSW	250	14/9.0 mJy
9 deg^2 (Pearson et al. 2017)	PMW	350	14.2/7.5 mJy
	PLW	500	13.8/10.8 mJy
SCUBA-2/NEPSC2[d]	850	850	1.0–2.3 mJy
2 deg^2 (Shim et al. 2020)			

Note. [a]The detection limits refer to the 4σ flux over a circular area with a diameter of 1 arcsec. [b]The detection limits refer to 3σ instrumental noise sensitivities. [c]The detection limits refer to the open time 2 (OT2) sensitivity. [d]The detection limits refer to the 1σ rms noise (or 4.7–11 mJy at 80% completeness).

82 per cent (75 000) sources in total, matched between the *AKARI* and HSC data without flagging. Within a 1.5 arcsec radius, 95 per cent of the sources (87 645) are matched.

We describe the details of the matching results in Fig. 4, which compares the distribution of subcategories of the sources divided into three (clean, flagged, and non-HSC) as a function of N2 magnitude. The grey histogram indicates all the N2 sources that correspond to the sum of the three (clean + flagged + non-HSC) groups. The violet histogram shows the N2 sources that have an HSC counterpart (clean + flagged). On the other hand, the green line shows the distribution of the N2 sources matched to the previous optical data (CFHT and/or Maidanak).

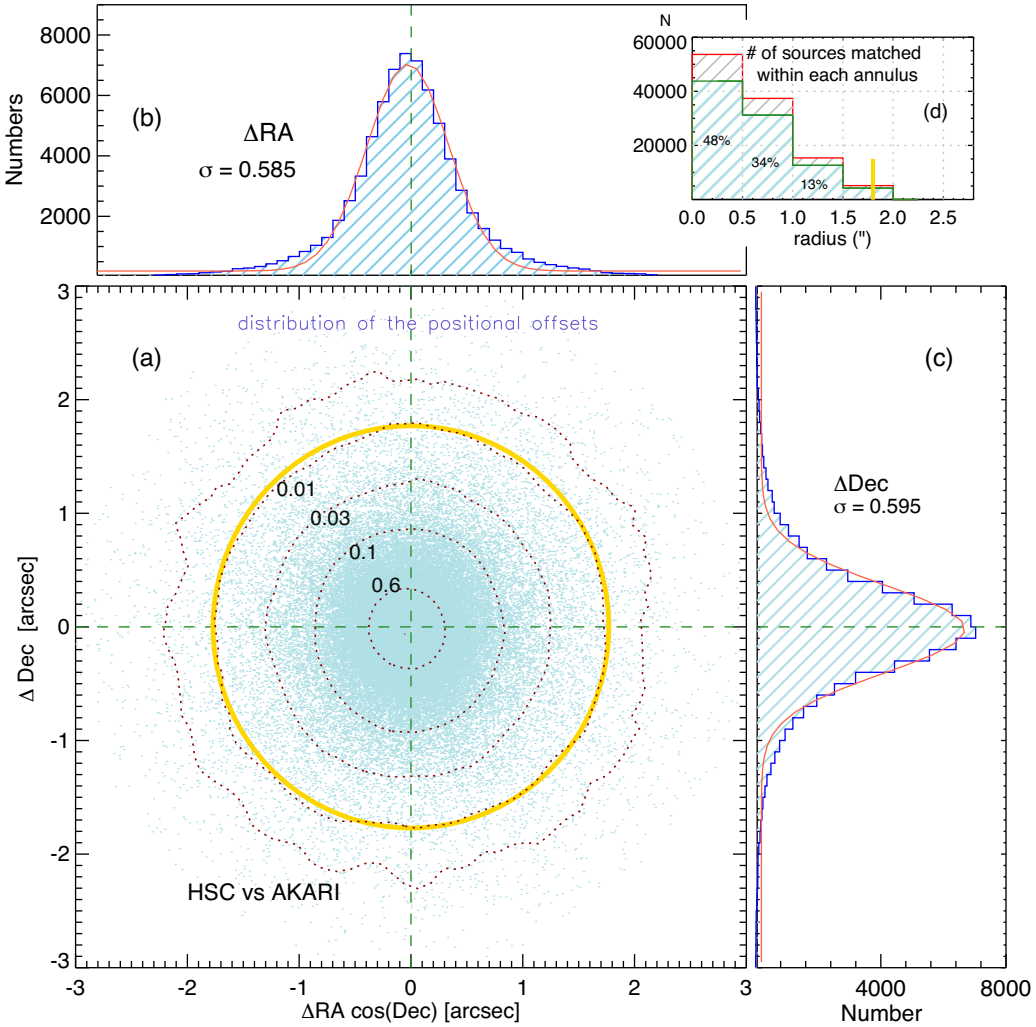

Figure 3. The positional offset distribution of the matched sources between the HSC and *AKARI* data. (a) The dotted contours with numbers represent the density levels normalized by the central peak density. The yellow circle is the matching radius determined based on the 3σ of the Gaussian, fitted to the histograms (magenta curves) in (b) and (c), shown in terms of a 0.2 arcsec bin. (d) Bars show how many sources were matched within the 0.5 arcsec annuli (the green bars indicate the number of clean sources, the red bars above the green ones show the increments by the flagged sources. See Section 2.3 for details).

In the bright magnitude range (up to 14.8 mag), there are no clean sources: all the sources are flagged (blue dotted) or non-HSC (red dashed), which means that most of the bright N2 sources either are accompanied by one of three HSC flags used to filter out problematic sources or do not have an HSC counterpart. This indicates that their HSC counterparts were affected by saturated/bad pixels, masked/edge regions, or do not have valid HSC parameters; otherwise they did not match any HSC sources. Between 13.5 and 17.5 mag, flagged sources prevail. Clean sources (black line) begin to appear around 15 mag and go above the flagged sources at 17.5 mag, and this predominance continues to a faint end. The N2 source counts decrease rapidly before the detection limit due to source confusion (Kim et al. 2012). We did not include sources fainter than the detection limit (i.e. a small number of objects fainter than the vertical dot–dashed line as shown in Fig. 4). Some of the non-HSC sources have optical counterparts in the previous CFHT/Maidanak data (in the brightest range) because they are located in the region where the *gizy* bands (which were used for the source detection) did not cover but the CFHT/Maidanak surveys provided normal photometric measurements (see the small bump in the red dashed histogram below the green one around 13–14 mag.). These sources are beyond the concern of this work and the *AKARI* sources classified as flagged and/or non-HSC might be discussed later (in separate works).

A similar description for the other *AKARI* bands is summarized in Fig. 5. The overall trends in the NIR bands are similar to N2: the red histogram in Fig. 5(a) for the N4 band shows a smaller bump around 14 mag and another peak around 20 mag. The smaller bump is ascribed to the unobserved area; otherwise they are probably the brightest stars (almost all of the NIR sources brighter than 15 mag seem to be stars; Kim et al. 2012), but are rejected by the pipeline or classified as flagged. Fig. 5(a) also shows there is no clean source brighter than 16 mag. This trend in the bright end weakens/disappears as we move to the mid-IR bands. It seems that the saturation levels of the HSC bands (17–18 mag, Aihara et al. 2018) correspond to the valley between the two red bumps in the NIR bands, where the clean sources begin to appear. The smaller red bump fades out in longer-wavelength (MIR) bands: it becomes weak in S7 (Fig. 5b), appears weaker in S9W (Fig. 5c), and completely disappears in the MIR-L bands, implying that the Rayleigh–Jeans tail of the stars fades out. In the NIR bands, the number of clean sources (black solid line) is higher than that of flagged sources (blue dotted) near the broad peak.

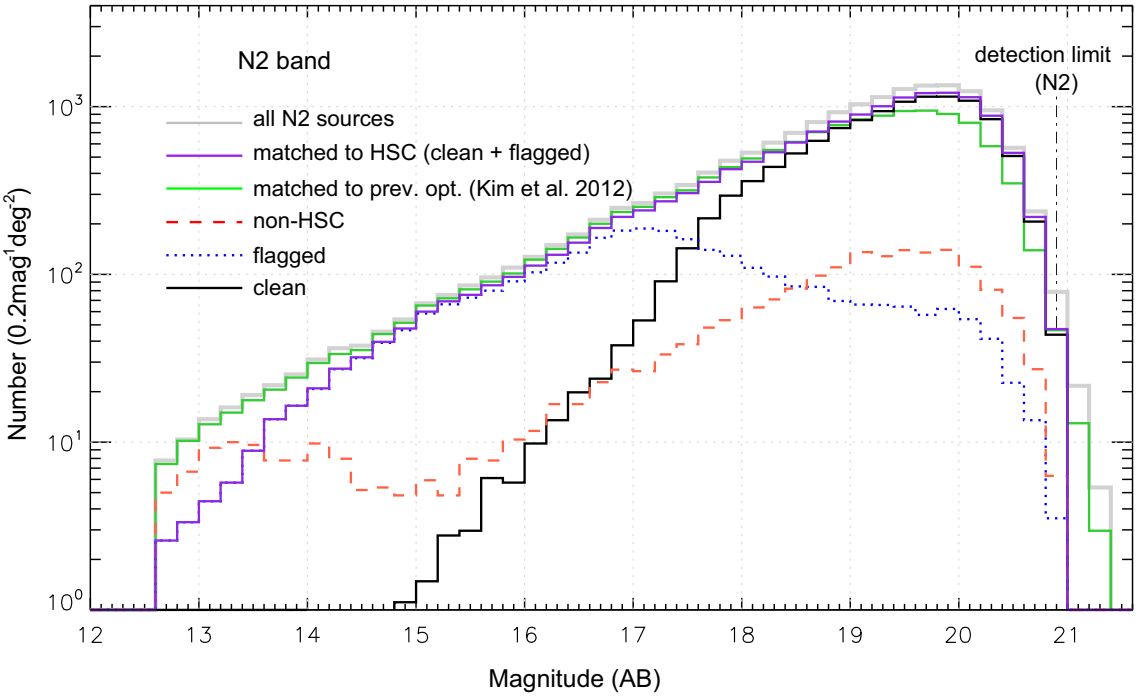

Figure 4. The number distribution of N2 sources as a function of magnitude (0.2 mag^{-1} deg^{-2}). All N2 sources are divided into three subcategories according to the matching results against the HSC data: clean (solid black), flagged (dotted blue), and non-HSC (dashed red). The violet line shows the optically matched sources (i.e. the sum of the clean and flagged sources). The grey line shows all the N2 sources (i.e. the sum of the clean, flagged, and non-HSC sources). The green line represents the N2 sources matched to the previous optical data from CFHT or Maidanak (Kim et al. 2012).

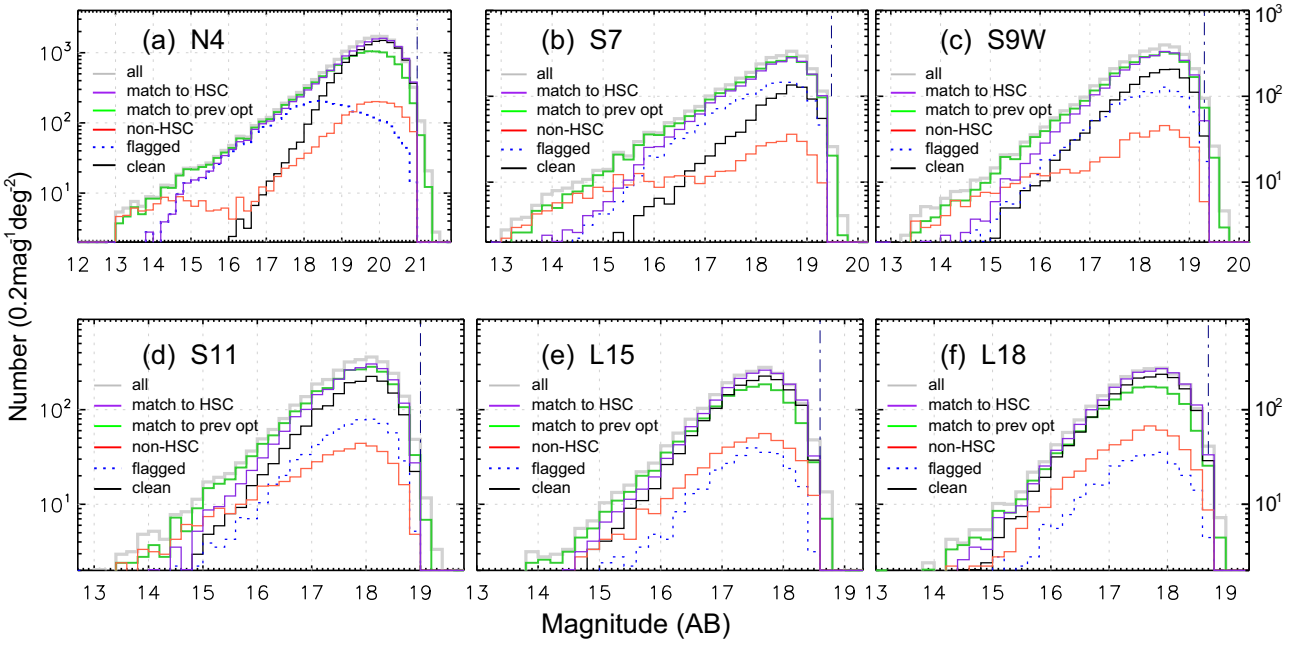

Figure 5. The number distribution of the sources as a function of magnitude, plotted in the same fashion as in Fig. 4, but for the sources in the other *AKARI* bands. The distributions for the three different subcategories are given in the same colours as shown in Fig. 4. The vertical dot–dashed line in each panel represents the detection limit in Table 1.

In the S9W bands, these two classes become comparable, while the number of non-HSC sources is much lower. In the S9W and S11 bands, a fraction of the flagged sources (blue dotted) are between the black and red lines. In the MIR-L bands, the blue dotted line becomes lower than non-HSC sources (red line).

Although some of the bright IR sources are not fully available in this work consequently because they are unobserved/rejected or eventually classified as flagged sources, on the other hand, there are many more fainter sources newly matched to the deeper HSC data, as shown in Fig. 6 (for example, 19 771 at N4 and 2593 at L15,

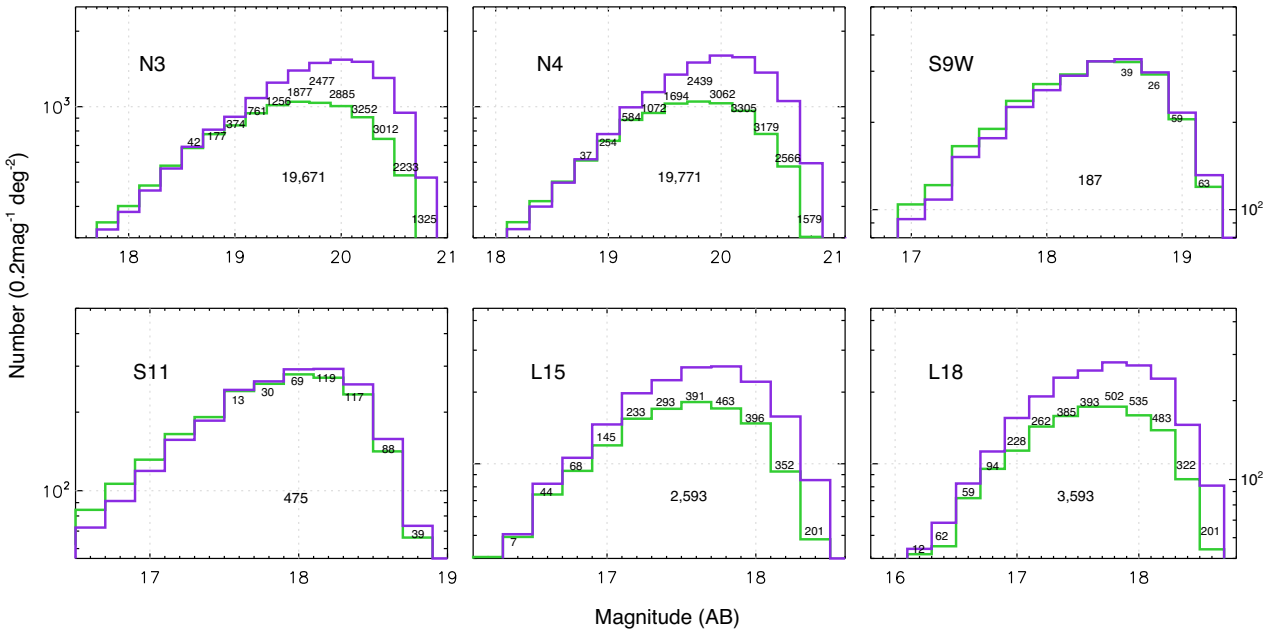

Figure 6. Selected narrow ranges from Fig. 5 near the peak of the histograms (to show the green and violet lines), which give the comparison between the sources matched to the HSC data (violet) and to the previous optical data obtained with CFHT and Maidanak (green). The small numbers written on the green histogram represent the difference between the violet and green histograms in each magnitude bin, i.e. newly matched to the HSC data. The sum of the small numbers is presented in the middle of each panel. In this range, the violet histogram being lower than green one (on the bright side) seems mostly due to the flagged sources (blue dotted line in Fig. 5) excluded from this work.

respectively). Only small ranges were presented in the figure: note the range where the violet histogram is higher than the green one. The number in each magnitude bin indicates the sources newly matched to the HSC data.

However, the bigger red bumps (in Fig. 5) near the faint ends indicate that our HSC survey was still not deep enough to identify all the faint IR sources, which left some of the *AKARI* sources unmatched against the HSC data. On the *AKARI* colour–colour diagrams (in Section 4), these faint IR sources are located in the same area as the other (optically identified) sources, which implies that they have similar IR properties. They are probably infrared luminous SFGs, but appear to have dropped out in the HSC bands. Seemingly, they might be a certain kind of highly obscured dusty system at high z. A detailed discussion based on the selected sample is presented in Toba et al. (2020).

3 COMPLEMENTARY DATA SETS

After the identification of *AKARI* IR sources with the HSC optical data, we used all available photometric catalogue/data over the NEPW field to construct a multiband catalogue. In this section, we briefly describe the data sets used in this work. Just as Fig. 1 showed the coverages of various surveys, Table 1 and Fig. 7 summarize the photometric bands and depths of the surveys. Fig. 7(a) also shows why source detection changes in different instruments/filter systems.

3.1 Ancillary optical data: CFHT and Maidanak

It is not easy to treat the entire 5.4 deg^2 area in a uniform manner unless we have a large-FoV instrument with an appropriate filter system covering a sufficient wavelength range. This was what divided our previous optical surveys into two different data sets 10 years ago: one obtained with the MegaCam (u^*, g, r, i, z) on the central 2 deg^2 area, and the other with the SNUCAM B, R, and I (Im et al. 2010) of the Maidanak observatory over the remaining part of the NEPW field. Detailed descriptions of these two surveys can be found in Hwang et al. (2007) and Jeon et al. (2010), respectively. However, the western half of the central CFHT field was not observed by the u^* band. Also, due to the different filter systems and depths between these two optical surveys, homogeneous analysis with optical counterparts over the whole field was practically impossible. Another optical survey was carried out later (Oi et al. 2014) on the NEPD field (∼0.7 deg^2) and finally provided MegaCam u^* data for the western-half area as well as the supplementary WIRcam data (Y, J, K_s). In addition, the CFHT MegaPrime u-band observation was performed over a 3.6 deg^2 area on the eastern side of the NEPW field (Huang et al. 2020)[5] to replenish the insufficient (central 2 deg^2 only; see Fig. 1) coverage of the MegaCam u^* band. Because the calibration of photo-z is a significant issue under circumstances in which a huge number of sources remain without redshift information, the availability of u-band data is crucial to check the UV extinction properties and to improve photo-z accuracy (Ho et al. 2020). We combined all these supplementary optical data: a small systematic shift of WCS (<1 arcsec) in each piece of optical data with respect to the HSC was corrected first, and the matching radii for each piece of data were decided based on the mean positional differences (see Fig. 8). The number of sources matched to the HSC data are summarized in Table 2.

3.2 Spectroscopic and photometric redshifts

Following the optical identification of the *AKARI* sources with the deep HSC data, we incorporated all available spectroscopic redshift

[5]http://doi.org/10.5281/zenodo.3980635

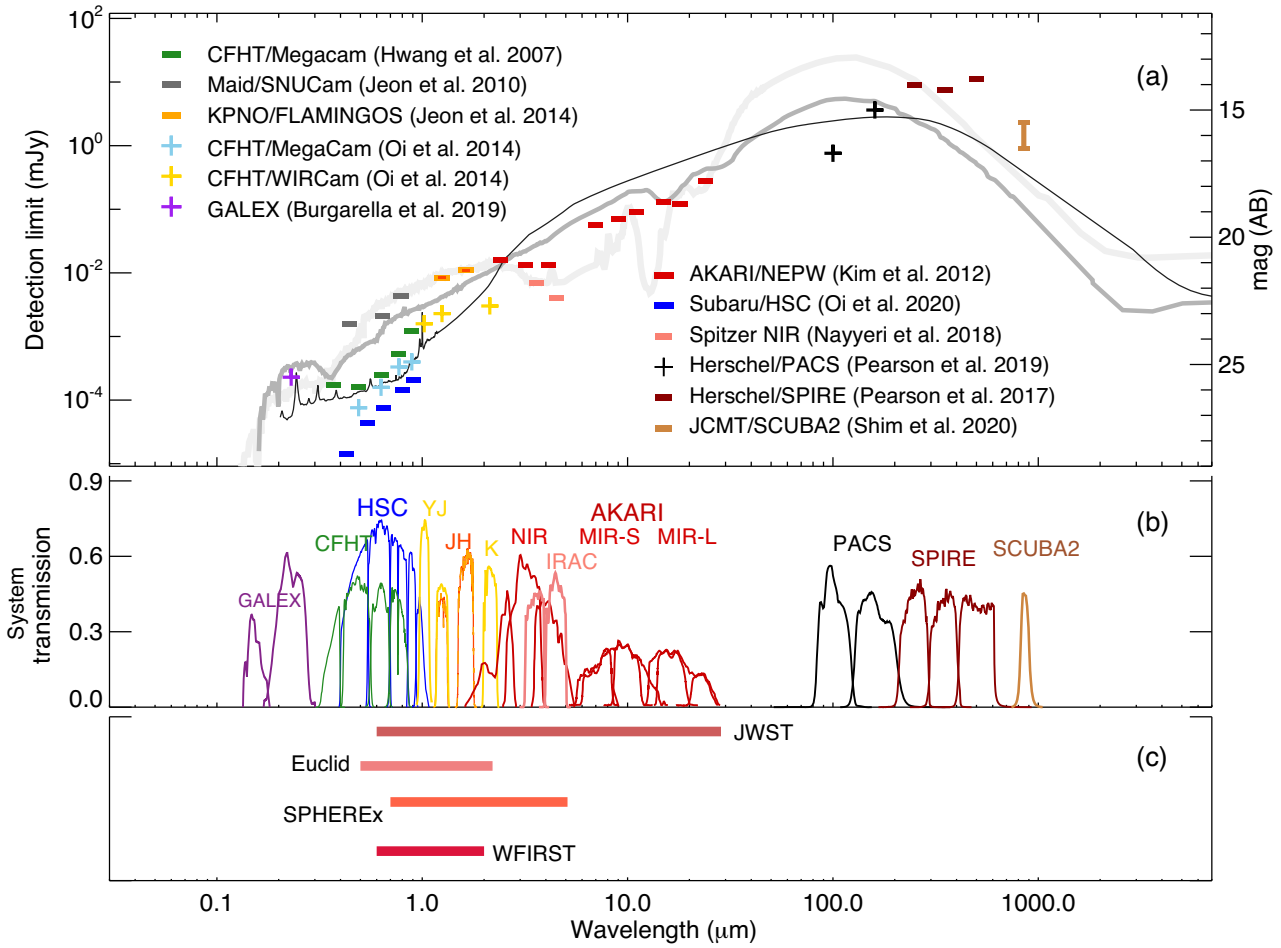

Figure 7. (a) The depths of various surveys with different instruments/filter systems over the NEP, in terms of the 5σ limiting magnitude. The cross symbols (+) imply that the survey was dedicated only to the NEPD field. Typical templates (Polletta et al. 2007) are given to show how a local ULIRG (e.g. Arp220-type at $z = 0.3$, faintest thick line), a type-1 Seyfert (at $z = 0.6$, dark grey), or a dusty torus model (at $z = 1.0$, thin black line) looks in this plot. All of them are normalized at the N2 detection limit. (b) The system transmission/filter shapes. (c) A comparison of the spectral range in the IR to be covered by future space missions, shown by horizontal bars.

(spec-z) data into the clean *AKARI*–HSC sources (therefore the redshift information matched to the flagged sources is not included here). There have been many spectroscopic observations over *AKARI*'s NEP area. The most extensive and massive campaign covering the entire NEPW field was done by Shim et al. (2013): they targeted the NEPW sources selected primarily based on the MIR fluxes at 11 μm (S11 < 18.5 mag) and at 15 μm (L15 < 17.9 mag) to see the properties of the MIR-selected SFGs. Most of these flux-limited sources turned out to be various types of IR luminous populations of galaxies. A smaller number of secondary targets (35 per cent of their targets) were also selected to catch some rare types of objects, such as obscured AGNs, BzKs, supercluster candidates, etc. They provided the spectra of 1796 sources (primary targets: 1155, secondary targets: 641), and the redshifts for 1645 sources were measured. These spectroscopic sources are classified into several types (e.g. star, type-1 AGN, type-2 AGN, galaxy, unknown). Recently, a new spectroscopic campaign over the whole NEPW area with the MMT/Hectospec has been initiated to carry out a homogeneous survey for the selected 9 μm galaxies (MMT2020A/B, PI: H. S. Hwang).

We also took redshift/type information from many other spectroscopic surveys on the NEPD field. For example, Keck/DEIMOS observations were conducted in order to measure the spectroscopic redshift and calibrate the photo-z for MIR galaxies (DEIMOS08) (Takagi et al. 2010), to measure [O II] luminosity against 8 μm luminosity (DEIMOS11) (Shogaki et al. 2018), and, more recently, to check the line emission evidence of AGNs and metallicity diagnostics of SFGs, etc. (DEIMOS14, DEIMOS15, DEIMOS17) (Kim et al. 2018). Another series of spectroscopic observations with Gran Telescope Canarias (GTC)/OSIRIS were carried out between 2014 and 2017 (e.g. GTC7-14AMEX, GTC4-15AMEX, GTC4-15BMEX, and GTC4-17MEX) (Díaz Tello et al. 2017) to see the X-ray signatures of highly obscured and/or Compton-thick (CT) AGNs along with the identification by the *Chandra* data (Krumpe et al. 2015). Subaru/FMOS spectroscopy was also obtained to investigate the mass–metallicity relation of IR SFGs in the NEP field (Oi et al. 2017). Ohyama et al. (2018) provided a polycyclic aromatic hydrocarbon (PAH) galaxy sample with redshift measurements through SPICY projects done by *AKARI*/slitless spectroscopy. We combined all this redshift information.

Using these spectroscopic redshifts as a calibration sample, Ho et al. (2020) estimated photo-z using photometry from the u^* band to the NIR band (IRAC2 and/or *WISE*2). They checked the photometry by comparing colours/magnitudes to discriminate unreasonable data so that they could obtain reliable results when using the software LE

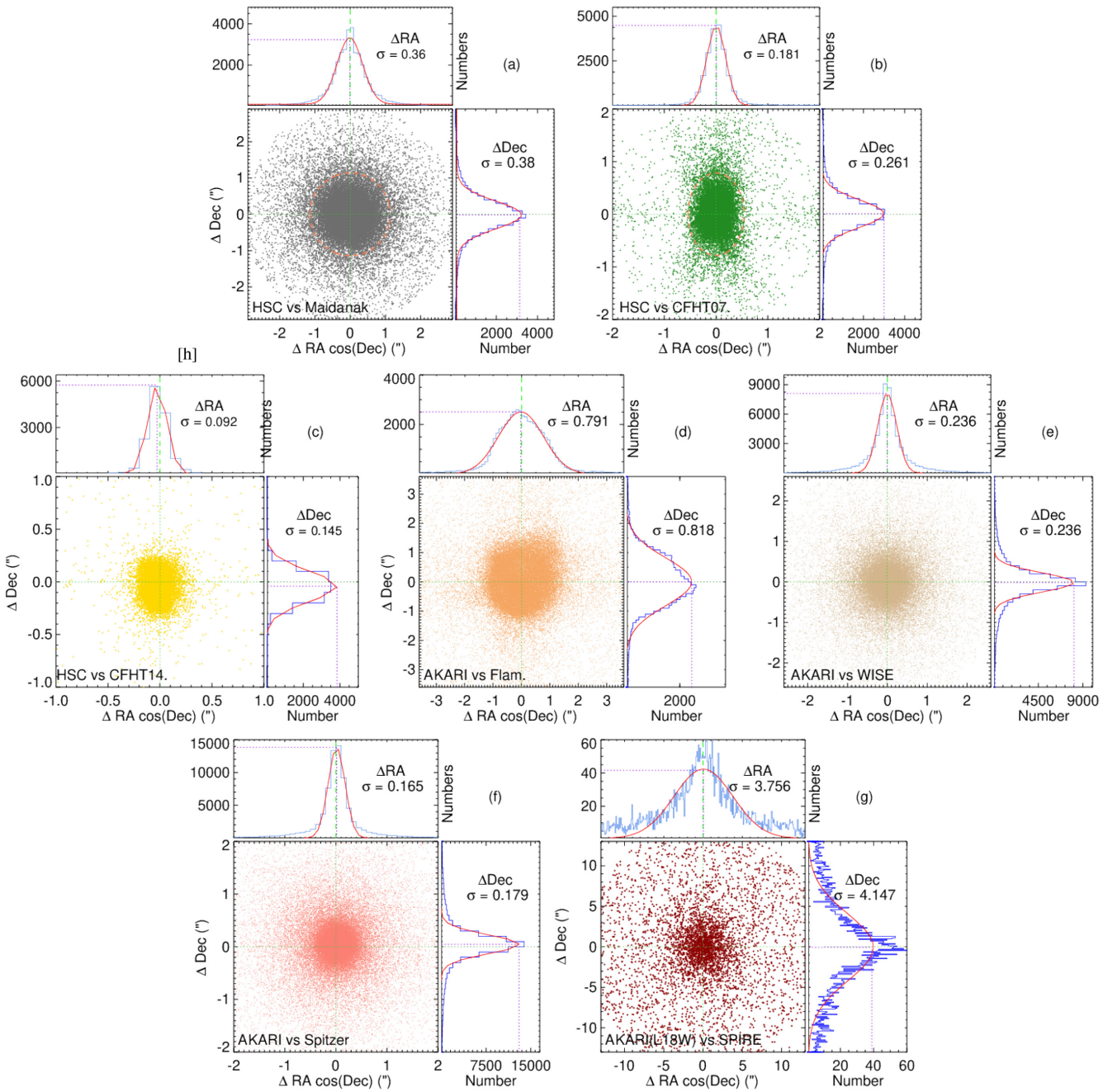

Figure 8. Some examples of the astrometry-offset distributions of the sources matched between the clean (HSC–*AKARI*) and the other supplementary data. The matching circles (or ellipses) were decided based on the representative value of the offset (width of the offset histogram), 3σ derived from the Gaussian fit to the histogram of ΔRA and ΔDec. Most of them appear to be circular except for the matching against the CFHT data. (a) HSC versus Maidanak, (b) HSC versus CFHT (2007), (c) HSC versus CFHT (2014), (d) *AKARI* versus FLAMINGOS, (e) *AKARI* versus *WISE*, (f) *AKARI* versus *Spitzer*, (g) *AKARI* (MIR-L) versus SPIRE.

PHARE. After the photo-z values were assigned, they presented an effective star–galaxy separation scheme based on the χ^2 values.

3.3 Supplementary near-/mid-IR data

While the optical data are crucial for identifying the nature of the corresponding *AKARI* sources, the *J*-, *H*-, and *K*-band data are useful to bridge the gap in wavelength between the optical *y* and N2 bands. The CFHT/WIRCam covered a limited area, but provided useful *J*- and *K*-band photometry (Oi et al. 2014). The NIR (*J*, *H*) survey that covered almost the entire NEPW area (5.2 deg^2) was done by FLAMINGOS of the Kitt Peak National Observatory (KPNO) 2.1 m telescope although the depth is shallower than the WIRcam data (see Figs 1 and 7). For complementary photometry in the near- to mid-IR, we also included the publicly available data taken by *Spitzer* (Werner et al. 2004) and *WISE*[6] (Wright et al. 2010). The catalogue by Nayyeri

[6]Also see https://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_1.html

Table 2. Summary of the matching against the supplementary data.

Main data	Supplementary data (reference)	3σ radius (arcsec)	PSF size (FWHM, arcsec)	Number of matched sources
HSC	Maidanak/SNUCam (Jeon et al. 2010)	1.14	1.1–1.4	33 485
Subaru	CFHT/MegaCam (Hwang et al. 2007)	0.54/0.78[a]	0.7–1.1	23 432
	CFHT/Mega-WIR (Oi et al. 2014)	0.28/0.44[a]	0.8–0.9	15 261
	CFHT/MegaPrime-u (Huang et al. 2020)	0.43/0.55[a]	0.8–1	31 851
	GALEX (Burgarella et al. 2019)	3.2	5.0	58
AKARI	KPNO/FLAMINGOS (Jeon et al. 2014)	2.2	1.7–1.8	46 544
	WISE (Jarrett et al. 2011)	0.9	∼6	60 062
	Spitzer (Nayyeri et al. 2018)	1.2	1.78	79 070
	PACS (Pearson et al. 2019)	3.6/6.3[b]	6.8/11.3	882/463
	SPIRE (Pearson et al. 2017)	8.1[c]	17.6	3109

Note. [a]The matching radii along the RA and Dec. are not the same (see Fig. 8). [b]The radii for the 100 and 160 μm bands, respectively. [c]The source extraction was done on the 250 μm map, and the sources were catalogued with photometry in all three SPIRE bands.

et al. (2018) provides 380 858 sources covering the entire NEPW field (∼7 deg^2) with higher sensitivity (21.9 and 22.4 mag at IRAC1 and IRAC2, respectively) and slightly better spatial resolution compared to N3 and N4, which are useful to cross-check against the longer-wavelength data with larger PSFs (e.g. the SPIRE or SCUBA-2 data).

3.4 FIR/SMM data from *Herschel* and SCUBA-2

Herschel (Pilbratt et al. 2010) carried out 0.44 and 9 deg^2 surveys over the NEP field with the Photoconductor Array Camera and Spectrometer (PACS, Poglitsch et al. 2010) and Spectral and Photometric Imaging REceiver instrument (SPIRE, Griffin et al. 2010), respectively. From the PACS NEP survey (Pearson et al. 2019; Burgarella et al. 2019), the green (100 μm) and red (160 μm) bands provide 1380 and 630 sources over the NEPD field, with flux densities of 6 and 19 mJy at the 50 per cent completeness level, respectively. SPIRE also carried out an NEP survey as an open time 2 programme (PI: S. Serjeant, in 2012), and completely covered the entire NEPW field, at 250, 350, and 500 μm (achieving 9.0, 7.5, and 10.8 mJy sensitivities at each band). Source extraction was carried out on the 250 μm map, and approximately 4800 sources were catalogued with the photometry in all three SPIRE bands. A more detailed description of the data reduction and photometry can be found in Pearson et al. (2017). Compared to the optical or NIR data, the *Herschel* (PACS or SPIRE) data have larger positional uncertainties with much larger PSF sizes. This can make the identification of sources against the *AKARI* data potentially ambiguous when we carry out positional matching, even though the radius was determined reasonably (the 3σ radii are smaller than the PSF sizes, in general, as shown in Table 2). In our catalogue, cases in which multiple *AKARI* clean sources are lying within the searching radius around the SPIRE/PACS positions were not included so that we clearly chose only one *AKARI* counterpart against the *Herschel* sources. The 850 μm submillimetre (sub-mm) mapping on the NEPW field is currently ongoing by one of the large programmes with the JCMT/SCUBA-2 (Shim et al. 2020). They released a mosaic map and a catalogue for the central 2 deg^2 area first. They provide 549 sources above 4σ with a depth of 1.0–2.3mJy per beam. The source matching against the *AKARI*–HSC clean catalogue was carried out based on the likelihood ratio. We derived the probability of a counterpart for a 850 μm source, using both the magnitude distributions of the IRAC1 and IRAC2 bands (which are deeper than those of the *AKARI* NIR bands) and their colour as well as those of three SPIRE bands. We took 46 sources as robust *AKARI* counterparts for the 850 μm sources because they are matched to both IRAC and SPIRE with high (95 per cent) probability. We also included 16 sources as decent *AKARI* counterparts because in these examples there is only one IRAC/SPIRE source within the 850μm beam. Lastly, four sources matched IRAC with high probability but we are uncertain about the SPIRE cross-identification. However, when multiple optical sources were associated with any given SPIRE or SCUBA-2 source, if a real optical counterpart was already classified as a flagged source, then it could be complicated or a potential issue.

4 THE PROPERTIES OF THE *AKARI* SOURCES IDENTIFIED BY HSC SURVEY DATA

4.1 Overview: reliability versus random matching

When we seek to find a counterpart against certain data using a searching radius (r), there is always a possibility that we could encounter a random source (which is not associated with a real counterpart) inside the circle defined by the radius (πr^2). This is the probability that a source can be captured simply by this circular area at any random position of the data. The higher the source density of data that we match against, the higher the random matching rate becomes. If we use a larger radius, then the probability is also increased. Kim et al. (2012) showed that it can be expressed in terms of the number density (n) of the data and the cross-section. See the open boxes and grey solid lines in Fig. 9(a), which are taken from fig. 14 in Kim et al. (2012).

In practice, these have to be regarded as upper limits of a false-match possibility (the worst case), because, in general, two data sets for source matching are correlated with each other: these data are obtained from the same field of view, not from two different arbitrary sky positions. Therefore, the matching results are generally better than this estimation (as indicated by the downward arrows).

To describe how to interpret the source matching in a way consistent with Kim et al. (2012), we used the same analyses, using our matching radii determined in Sections 2 and 3, and compared our results with the plots in the previous work. In Kim et al. (2012), they used the uniform radius (3 arcsec, the open boxes in Fig. 9a). In this work, however, the matching radii were chosen based on the mean positional offset of the matched sources between the data. This gives much smaller radii compared to those from Kim

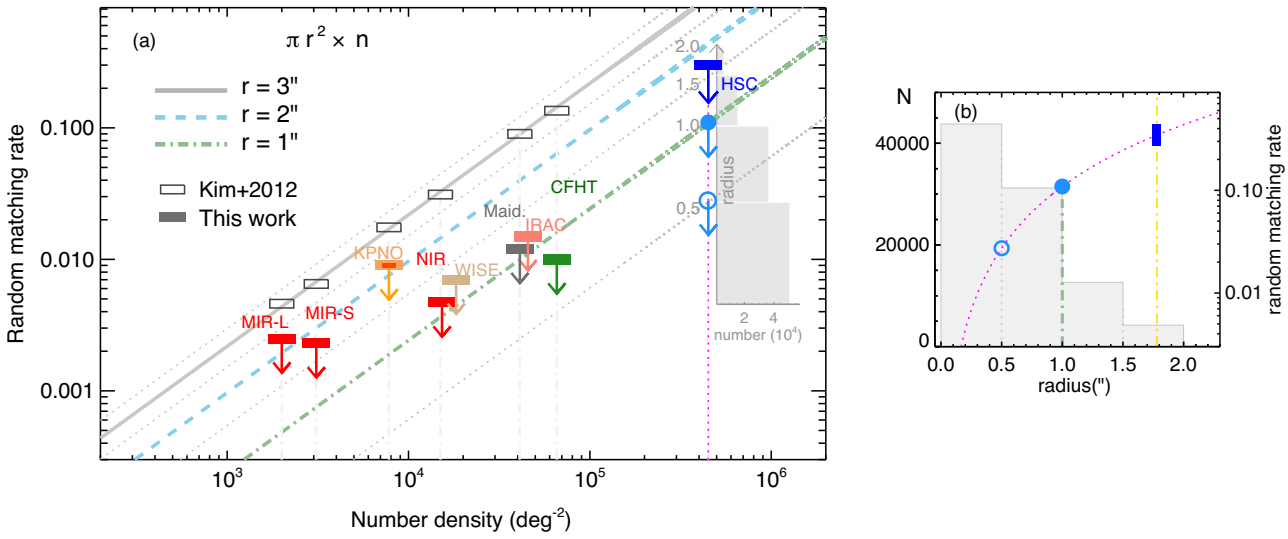

Figure 9. (a) Random matching rate in terms of the source density and matching radius. The open boxes are taken from fig. 14 in Kim et al. (2012) and represent the actual tests, i.e. number counts of the sources randomly matched to each piece of data using a 3 arcsec radius. The thick grey line shows that these measurements are described by a simple relation ($n\pi r^2$). If we use a radius of 2 or 1 arcsec, the random matching rate will decrease as described by the cyan dashed and green dot–dashed lines. The faint dotted lines between them indicate 0.5 arcsec increments. The filled boxes in different colours show the random matching estimates when a certain source is matched against the data with our radii determined in this work. The grey histogram is taken from Fig. 3(d). (b) The embedded grey histogram in (a) can be plotted with random matching rate as shown. The magenta curve shows the random matching rate when the number density of the source is 4.5×10^5 (which corresponds to the magenta vertical line in panel a). The vertical yellow line represents the matching radius from Fig. 3.

et al. (2012), even smaller than the PSF sizes, which are also used frequently for matching criteria. In Fig. 9(a), the highest number density of the HSC data ($\sim 4.5 \times 10^5$ sources per deg^2) seems to have a high random matching (the filled blue box). Here, we should remind ourselves that only 5 per cent of the sources are matched outside 1.5 arcsec and 82 per cent of the sources are matched within 1 arcsec, as shown by the grey background histogram, which is the same as Fig. 3(d) (but the *x*- and *y*-axes are transposed/rotated here). Therefore, the random matching rate is better than the filled circle on the green dot–dashed line. In Fig. 9(b), this grey histogram is replotted with a random matching rate, which gives a more straightforward description. However, it should be noted that this is just an estimation of the probability when the test is performed on random positions. Only a small fraction will suffer from random matching, in reality. The sources matched within 0.5 arcsec seem obviously reliable (downward below the open circle). In the same fashion, the matching with the other supplementary data (in green, grey, salmon boxes, and so on), which have a much lower number density, are less affected by random matching and are relatively safe compared to the HSC data.

4.2 Colour–colour diagrams

We describe the photometric properties of the NEPW infrared sources matched to the HSC optical data using various colour–colour diagrams (Figs 10–12). The colour–colour diagrams are also helpful to see from a statistical standpoint, if the source matching was accomplished properly. In each diagram, we used several different colours and symbols to distinguish between the different types of sources. Violet dots indicate the sources classified as star-like, which were fitted better to the stellar-model templates rather than the galaxy templates, following the diagnostics ($\chi^2_{\rm star} < \chi^2_{\rm galaxy}$) by Ilbert et al. (2009) when the photo-*z* estimation was performed with LE PHARE (Ho et al. 2020). The sources fitted better to the galaxy templates are divided into two groups by redshift: black dots represent the local ($z_{\rm phot} < 1$) galaxies and the grey dots represent the high-*z* (>1) objects. To see if the star-like sources are classified appropriately, we over-plotted (in cyan) the sources having high stellarity (i.e. star–galaxy classifier; sgc>0.85), measured with the SExtractor on the CFHT data (Hwang et al. 2007; Jeon et al. 2010). The stellar sequence appears prominently in the optical colour space because the stellar SED shows blackbody-like behaviour, which naturally generates a systematically and statistically well defined sequence.

In our spectroscopic sample (as explained in Section 3.2), we have various types classified by the line emission diagnostics. We have also plotted some of them here: spectroscopically confirmed stars (marked by yellow stars), type-1 AGNs (red crosses), AGNs identified in X-ray data (green squares), and sub-mm galaxies detected in the SCUBA-2 survey (salmon diamonds).

Fig. 10 shows the colour–colour diagrams using the photometry in the HSC and *AKARI* NIR bands. The violet dots form a very distinct track, cyan dots are exactly tracing this track, and five spectrocopic stars are overlaid on them. Those star-like sources (as well as the point sources) seem to be representative of the Galactic stars, but it is obvious that not all of them are stars (i.e. quite a few real galaxies, which just happen to fall on the stellar locus, are included in the vicinity of the sequence). This implies that the source matching was properly achieved and the star–galaxy separation was also effectively done. In the optical colour–colour diagrams, the stellar sequence overlaps with extragalactic sources, positionally entangled with them on the same area (Figs 10a and b), but when the NIR bands are involved, this stellar sequence gets separated from the extra-galactic populations (Figs 10c and d). In the NIR colour–colour diagrams, however, stars seem to stay together in a rather circular/elliptical area, not in the form of a track (Fig. 11). They gradually disappear in the longer wavelengths (Fig. 12).

In the optical colour–colour diagrams (Fig. 10), the black and grey dots (local and high-*z* galaxies, respectively) are gathered mostly in the same area, but the grey dots seem to be more widespread.

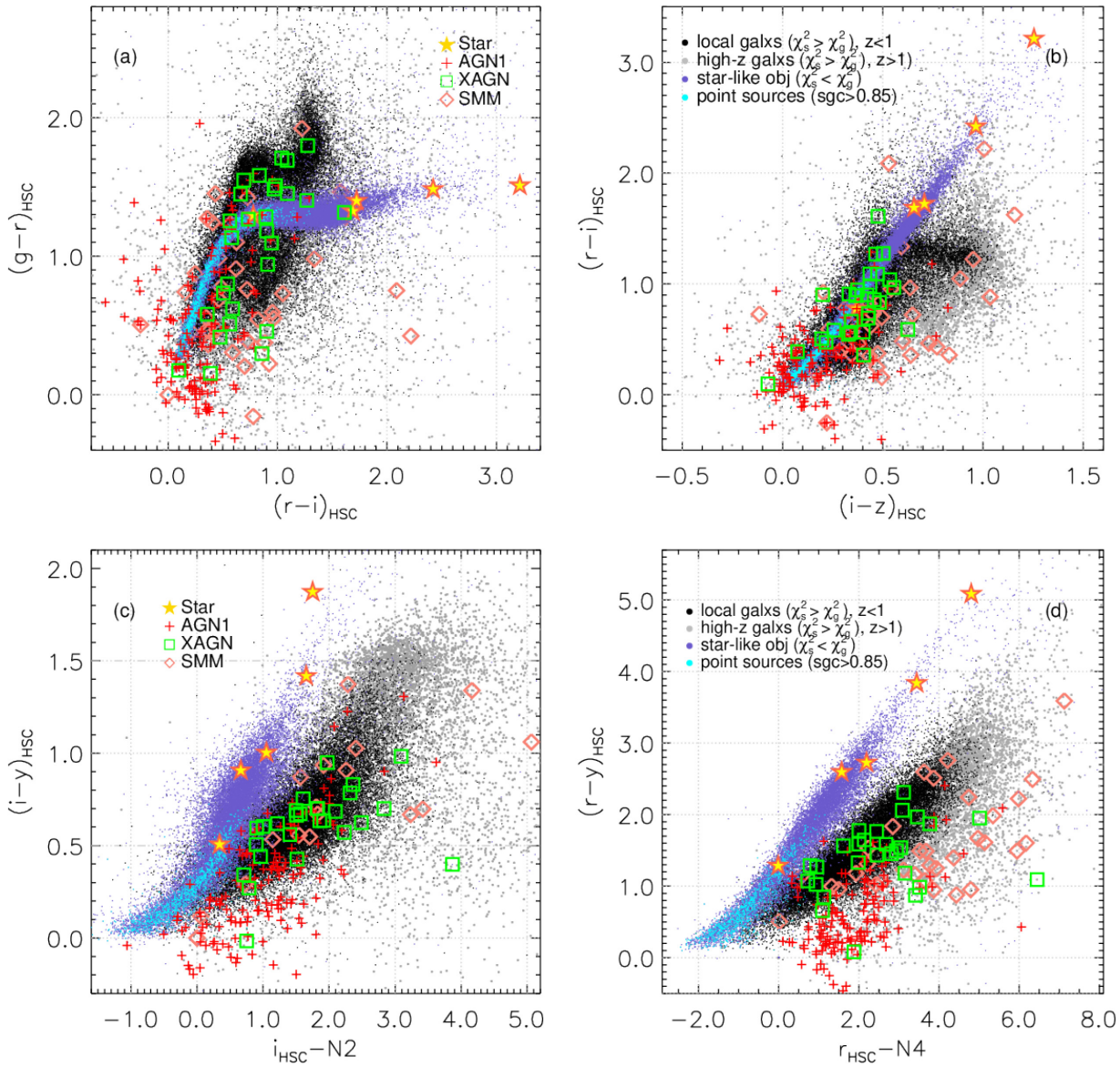

Figure 10. Colour–colour diagrams based on the HSC optical and *AKARI* NIR bands. Violet dots represent the sources classified as star-like sources; black dots represent the extra-galactic sources with $z_{phot} < 1$ while the grey dots are sources with $z_{phot} > 1$ (Ho et al. 2020). Cyan dots are high-stellarity (sgc>0.8) sources. Yellow stars represent the Galactic stars observed by the spectroscopic survey (Shim et al. 2013). Red crosses are AGNs (type 1), also confirmed by Shim et al. (2013). Green boxes are AGNs that have X-ray data (Krumpe et al. 2015; Díaz Tello et al. 2017). Salmon diamonds are galaxies observed by the SCUBA-2 survey (Shim et al. 2020). All axes are in units of AB magnitude.

In Fig. 11, the black dots seem to be gathering in an apparently different place from the grey dots which are spread towards the redder direction, being consistent with the photo-z classification and implying high-z populations. This separation becomes obvious in N2 − N4 colour (Fig. 12b and 12c), which appears to be a very good selector of high-z objects or AGNs, seemingly related to hot dust heated by energetic sources (Fritz, Franceschini & Hatziminaoglou 2006; Lee et al. 2009).

In Figs 11 and 12, we present the redshift tracks ($0 < z < 5.5$) to compare with the colour–colour diagrams based on the HSC and *AKARI* NIR bands (see Figs 11a′, b′ and 12a′, b′). They enable us to obtain the characteristics of the sources by comparing the trajectories of typical models with real galaxies observed by optical–IR surveys (as well as the symbols in Figs 11a, b and 12a, b). In Fig. 12, it should be noted that the model tracks show some overlaps in a certain area, suggesting that there seems to be partly a mixture of SFGs and AGNs. To select AGN types, the N2 − N4 colour seems to be more effective when we use a combination with MIR bands (e.g. N4 − S7 or S7 − S11).

The AGN types and SMM galaxies stay close to the black/grey dots. It is not easy to discern clear trends. However, the green boxes are widely overlapped with all the extragalacic sources (black and grey dots), while the red crosses (type 1) tend to stay in a specific area, all through the colour–colour diagrams. Salmon diamonds (SMM galaxies) are also spread over the black and grey dots, but more widely spread compared to the green boxes and they seem to prefer to stay around the grey dots, implying that the SMM galaxies are more likely to be high-z populations. On the other hand, it would be interesting to see some follow-up studies (e.g. Poliszczuk et al., in preparation; Chen et al., in preparation) on whether machine-

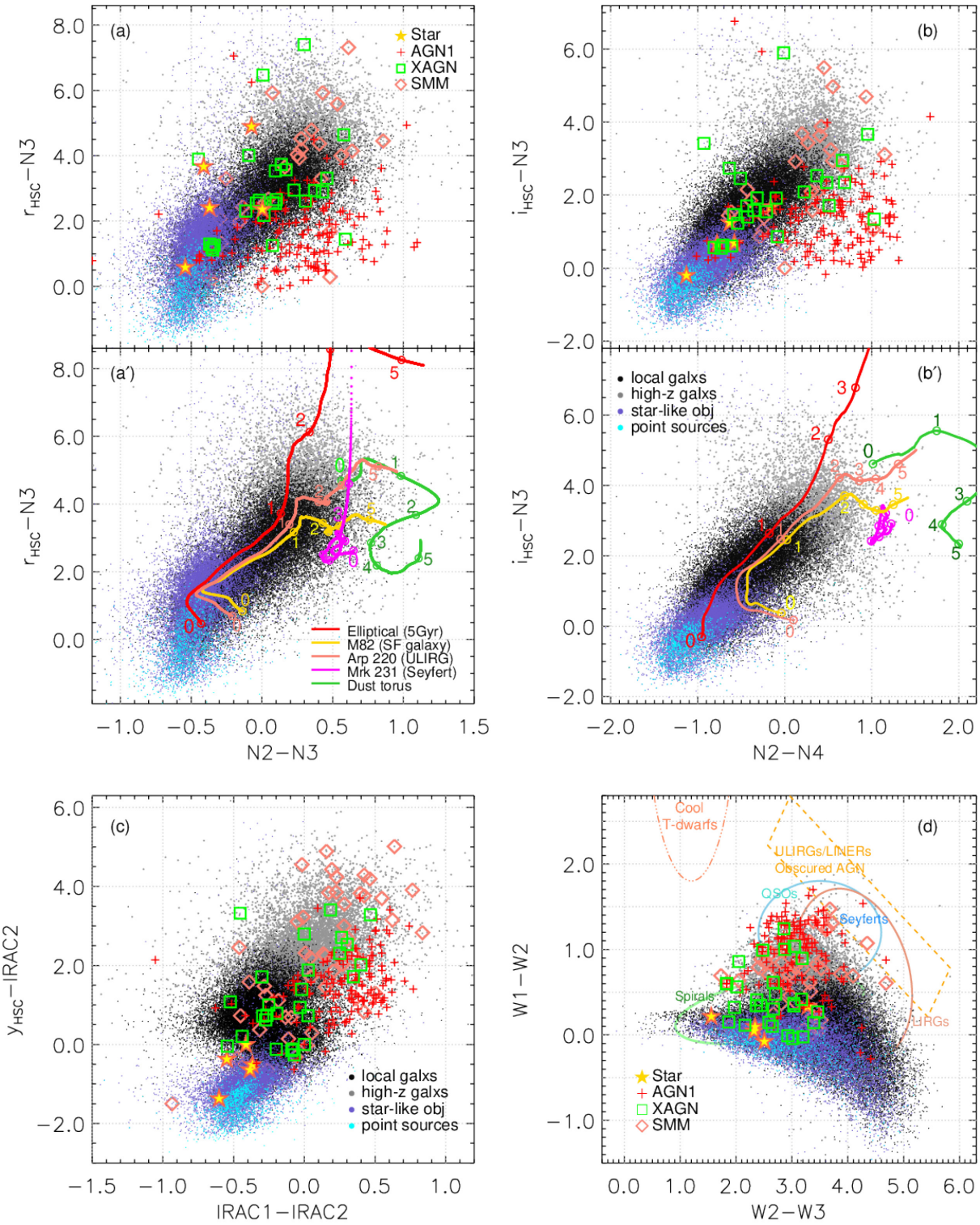

Figure 11. Colour–colour diagrams based on the HSC and various NIR bands (*AKARI*, *Spitzer*, and *WISE*), with the symbols in the same fashion as presented in Fig. 10: star–galaxy separation and how their locations change in different colour–colour diagrams are described. In (a′) and (b′), we present the evolutionary tracks of several model templates from Polletta et al. (2007). All axes are in units of AB magnitude except for (d), which is given in units of Vega magnitude to compare with the diagram from Wright et al. (2010). The magnitude offsets (Δm) between the AB and Vega systems ($m_{AB} = m_{Vega} + \Delta m$) are 2.70, 3.34, and 5.17 mag for W1, W2, and W3, respectively.

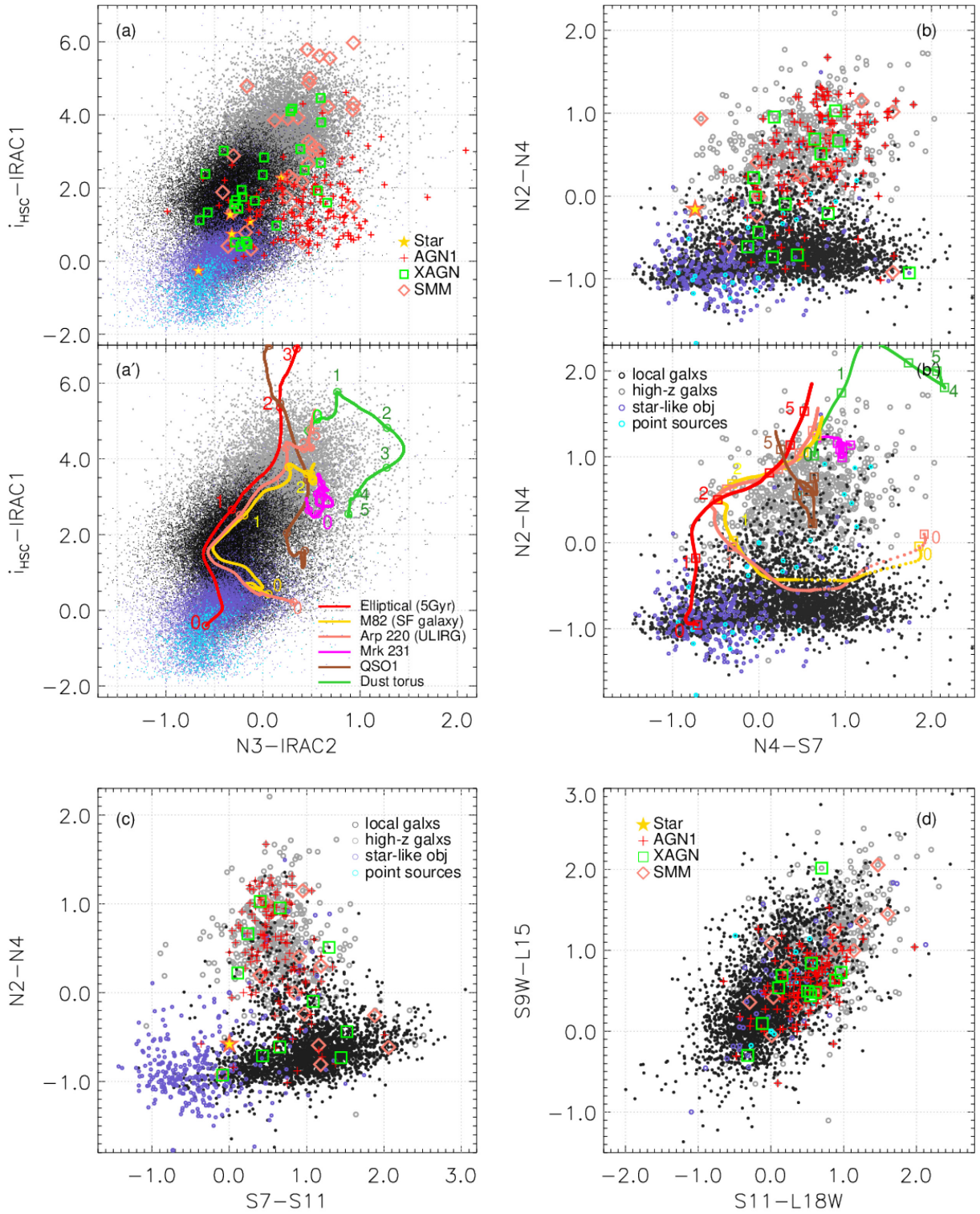

Figure 12. Colour–colour diagrams based on the various combination of NIR to MIR bands, in the same fashion as presented in Figs 10 and 11. Stars begin to fade out in MIR bands. In (a) and (b), we also present the evolutionary tracks (a′ and b′) of several model templates from Polletta et al. (2007). All axes are in units of AB magnitude.

learning algorithms, such as the support vector machine or deep neural network, etc., can do more effective separations in various colour/parameter spaces (not just in two-dimensional projections of them).

5 SUMMARY AND CONCLUSION

The NEP field has been a longstanding target since it was surveyed by the legacy programme of the *AKARI* space telescope (Serjeant et al. 2012). Previous optical surveys (Hwang et al. 2007; Jeon et al.

2010) were incomplete, which was a strong movitation to obtain deep Subaru/HSC optical data covering the entire NEP field (Goto et al. 2017). We have achieved faint detection limits (Oi et al. 2020), which enabled us to identify faint *AKARI* sources in the near- and mid-IR bands, and initiated a variety of new studies. We have constructed a band-merged catalogue containing photometric information for 42 bands from UV to the sub-mm 500 μm. The photo-z values for the NEPW sources were derived based on these data with all available redshift information (Ho et al. 2020) and were incorporated into the catalogue as well. We investigated the photometric properties of the NEPW sources observed by the HSC using colour–colour diagrams based on this band-merged catalogue. We are able to roughly see how the shape of the stellar sequence changes and the areas in which the AGN types prefer to stay in different colour spaces as the observed wavelength increases, although it is difficult to draw conclusions about the clear trend of extragalactic populations because a quantitative analysis has not been done.

This band merging gives us the benefits of constructing full SEDs for abundant dusty galaxy samples for SED modelling, e.g. using CIGALE (Boquien et al. 2019) or MAGPHYS (da Cunha, Charlot & Elbaz 2008), especially taking advantage of the uniqueness of the continuous MIR coverage as well as a wide range of panchromatic photometry. It provides more opportunities to disentangle otherwise degenerate properties of galaxies or to excavate hidden information for a better understanding of the physics behind various IR features.

Due to the uniqueness of the filter coverage by *AKARI*, these legacy data remain the only survey having continuous mid-IR imaging until *JWST* carries out its first look at the sky (e.g., Jansen and Windhorst, 2018). The science on this NEP field is currently driven by Subaru/HSC (Oi et al. 2020), SCUBA-2 (Shim et al. 2020), and homogeneous spectroscopic surveys. Since many future space missions are planning to conduct deep observations of this area, e.g. *Euclid* (Laureijs et al. 2011), *JWST* (Gardner et al. 2006), *SPHEREx* (Doré et al. 2016, 2018), etc., a great deal of synergy is expected together with the legacy data as well as our ongoing campaigns.

ACKNOWLEDGEMENTS

We thank the referees for the careful reading and constructive suggestions to improve this paper. This work is based on observations with *AKARI*, a JAXA project with the participation of ESA, universities, and companies in Japan, Korea, and the UK. This work is based on observations made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech. *Herschel* is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA. TG acknowledges the support from the Ministry of Science and Technology of Taiwan through grants 105-2112-M-007-003-MY3 and 108-2628-M-007-004-MY3. This research was conducted under the agreement on scientific cooperation between the Polish Academy of Sciences and the Ministry of Science and Technology in Taipei and supported by the Polish National Science Centre grant UMO-2018/30/M/ST9/00757. HS acknowledges the support from the National Research Foundation of Korea grant No. 2018R1C1B6008498. TH is supported by the Centre for Informatics and Computation in Astronomy (CICA) at National Tsing Hua University (NTHU) through a grant from the Ministry of Education of the Republic of China (Taiwan).

DATA AVAILABILITY

The band-merged catalogue in this work is available at Zenodo (https://zenodo.org/record/4007668#.X5aG8XX7SuQ). Other data addressed in this work will be shared on reasonable request to the corresponding author.

REFERENCES

Aihara H. et al., 2018, PASJ, 70, S4
Barrufet L. et al., 2020, A&A, 641, A129
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Boquien M. et al., 2019, A&A, 622, A103
Bosch J. et al., 2018, PASJ, 70, S5
Burgarella D. et al., 2019, PASJ, 71, 12
Casey C. M., Narayanan D., Cooray A., 2014, Phys. Rep., 541, 45
Chiang C.-Y. et al., 2019, PASJ, 71, 31
da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595
Doré O. et al., 2016, preprint (arXiv:1606.07039)
Doré O. et al., 2018, preprint (arXiv:1805.05489)
Díaz Tello J. et al., 2017, A&A, 604, A14
Fritz J., Franceschini A., Hatziminaoglou E., 2006, MNRAS, 366, 767
Galliano F., Galametz M., Jones A. P., 2018, ARA&A, 56, 625
Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485
Geach J. E. et al., 2017, MNRAS, 465, 1789
Goto T. et al., 2019, PASJ, 71, 30
Goto T. et al., 2010, MNRAS, 514, A6
Goto T. et al., 2017, Publ. Korean Astron. Soc., 32, 225
Griffin M. J. et al., 2010, A&A, 518, L3
Henry J. P., Mullis C. R., Voges W., Böhringer H., Briel U., Gioia I. M., Huchra J. P., 2006, ApJS, 162, 304
Hickox R. C., Alexander D. M., 2018, ARA&A, 56, 625
Ho C.-C. et al., 2020, MNRAS, in press
Holloway J., 1986, J. Appl. Phys., 60, 1091
Huang S. et al., 2018, PASJ, 70, S6 ,
Huang T.-C. et al., 2020, MNRAS, 498, 609
Hwang H. S. et al., 2010, MNRAS, 409, 75
Hwang H. S., Geller M. J., Kurtz M. J., Dell'Antonio, I. P., Fabricant, D. G., 2012, A&A, 758, 25
Hwang H. S., Serjeant S., Lee M. G., Lee K. H., White G. J., 2007, MNRAS, 375, 115
Hwang N. et al., 2007, ApJS, 172, 583
Ilbert O. et al., 2009, ApJ, 690, 1236
Im M., Ko J., Cho Y., Choi C., Jeon Y., Lee I., Ibrahimov M., 2010, J. Korean Astron. Soc., 43, 75
Jansen R. A., Windhorst R. A., 2018, PASP, 130, 124001
Jarrett T. H. et al., 2011, ApJ, 735, 112
Jeon Y., Im M., Ibrahimov M., Lee H. M., Lee I., Lee M. G., 2010, ApJS, 190, 166
Jeon Y., Im M., Kang E., Lee H. M., Matsuhara H., 2014, ApJS, 214, 20
Kim H. K. et al., 2018, in Ootsubo T., Yamamura I., Murata K., Onaka T., The Cosmic Wheel and the Legacy of the AKARI Archive: From Galaxies and Stars to Planets and Life. p. 371
Kim S. J. et al., 2019, PASJ, 71, 11
Kim S. J. et al., 2012, A&A, 548, A29
KirkpatrickA. et al., 2012, ApJ, 759, 139
Kumpe M. et al., 2015, MNRAS, 446, 911
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lee H. M. et al., 2009, PASJ, 61, 375
Lutz D., 2014, ARA&A, 52, 373
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Matsuhara H. et al., 2006, PASJ, 58, 673
Merloni A. et al., 2012, preprint (arXiv:1209.3114)
Miyazaki S. et al., 2018, PASJ, 70, S1
Murakami H. et al., 2007, PASJ, 59, S369
Nayyeri H. et al., 2018, ApJS, 234, 38
Offenberg J. D. et al., 2001, PASP, 113, 240

Ohyama Y. et al., 2018, A&A, 618, A101
Oi N., Goto T., Malkan M., Pearson C., Matsuhara H., 2017, PASJ, 69, 70
Oi N. et al., 2020, MNRAS, in press
Oi N. et al., 2014, A&A, 566, A60
Onaka T. et al., 2007, PASJ, 59, S401
Pearson C. et al., 2019, PASJ, 71, 13
Pearson C. et al., 2017, Publ. Korean Astron. Soc., 32, 219
Pilbratt G. L. et al., 2010, A&A, 518, L1
Poglitsch A. et al., 2010, A&A, 518, L2
Polletta M. et al., 2007, ApJ, 663, 81
Sanders D. B., 2014, Advances Space Res., 34, 535
Serjeant S. et al., 2012, preprint (arXiv:1209.3790)
Shim H. et al., 2013, ApJS, 207, 37
Shim H.et al., 2020, MNRAS, 498, 5065
Shogaki A. et al., 2018, in Ootsubo T., Yamamura I., Murata K. Onaka T., eds, The Cosmic Wheel and the Legacy of the AKARI Archive: From Galaxies and Stars to Planets and Life. p. 367
Spergel D. et al., 2015, preprint (arXiv:1503.03757)
Sutherland W., Saunders W., 1992, MNRAS, 259, 413
Takagi T. et al., 2010, A&A, 514, A5
Takagi T. et al., 2012, A&A, 537, A24
Toba Y. et al., 2015, PASJ, 67, 86
Toba Y. et al., 2020, ApJ, 899, 35
Wada T. et al., 2008, PASJ, 60, S517
Wang T.-W. et al., 2020, MNRAS, 499, 4068
Werner M. W. et al., 2004, ApJS, 154, 1
White G. J. et al., 2010, A&A, 517, A54
White G. J. et al., 2017, Publ. Korean Astron. Soc., 32, 231
Wright E. L. et al., 2010, AJ, 140, 1868

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

[1]*Institute of Astronomy, National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu 30013, Taiwan*
[2]*Faculty of Science Division II, Liberal Arts, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku, Tokyo 162-8601, Japan*
[3]*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[4]*National Institute of Technology, Wakayama College, 77 Noshima, Nada-cho, Gobo, Wakayama 644-0023, Japan*
[5]*Department of Earth Science Education, Kyungpook National University, Daegu 41566, Korea*
[6]*Department of Astronomy, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan*
[7]*Academia Sinica Institute of Astronomy and Astrophysics, 11F Astronomy-Mathematics Building, AS/NTU, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan*
[8]*Research Center for Space and Cosmic Evolution, Ehime University, 2-5 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan*
[9]*Korea Astronomy and Space Science Institute (KASI), 776 Daedeok-daero, Yuseong-gu, Daejeon 34055, Korea*
[10]*Centre for Informatics and Computation in Astronomy (CICA), National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu 30013, Taiwan (ROC)*
[11]*European Space Astronoy Center, E-28691 Villanueva de la Canãda, Spain*
[12]*RAL Space, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK*
[13]*Department of Physics and Astronomy, UCLA, 475 Portola Plaza, Los Angeles, CA 90095-1547, USA*
[14]*Department of Space and Astronautical Science, Graduate University for Advanced Studies, SOKENDAI, Hayama, Miura District, Kanagawa 240-0193, Japan*
[15]*Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, 252-5210 Kanagawa, Japan*
[16]*Instituto de Astronomía, Universidad Nacional Autónoma de México (UNAM), Km. 107, Carret. Tij.-Ens., Ensenada 22860, Mexico*
[17]*The Open University, Milton Keynes MK7 6AA, UK*
[18]*Oxford Astrophysics, University of Oxford, Keble Rd, Oxford OX1 3RH, UK*
[19]*National Centre for Nuclear Research, ul. Pasteura 7, PL-02-093 Warsaw, Poland*
[20]*Astronomical Observatory of the Jagiellonian University, ul. Orla 171, PL-30-244 Cracow, Poland*
[21]*Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ky, Tokyo 113-0033, Japan*
[22] *Japan Space Forum, 3-2-1 Kandasurugadai, Chiyoda-ku, Tokyo 101-0062, Japan*

This paper has been typeset from a TEX/LATEX file prepared by the author.